Psychological Monographs
General and Applied

o. 512
1961

Hilgard, Weitzenhoffer, Landes, and Moore

Vol. 75
No. 8

The Distribution of Susceptibility to Hypnosis in a Student Population:

A Study Using the Stanford Hypnotic Susceptibility Scale

By

Ernest R. Hilgard, André M. Weitzenhoffer
Judah Landes, and Rosemarie K. Moore

Stanford University

Price $1.00

Edited by Norman L. Munn
Published by the American Psychological Association, Inc.

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology* with the *Psychological Monographs*

NORMAN L. MUNN, Editor

Department of Psychology, Bowdoin College
Brunswick, Maine

Manuscripts and correspondence on editorial matters should be sent to the Editor. *Psychological Monographs* publishes comprehensive experimental investigations and other psychological studies which do not lend themselves to adequate presentation as journal articles. Major space is given to the author's original contribution; introductory and bibliographic materials, as well as statistical tables and graphs, must be kept within reasonable bounds. Tables, graphs, and appendix materials which deal with detail not essential to adequate presentation of the findings may be made available through the American Documentation Institute—for details of this procedure, see the APA *Publication Manual*. Preparation of manuscripts for publication as monographs should follow the procedure given in the APA *Publication Manual*. Publication in *Psychological Monographs* is free of cost to the author, except in cases where early publication is requested or author's alterations are made in galley proofs.

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

Correspondence on business matters should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St., N.W., Washington 6, D.C. Address changes must arrive by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee third-class forwarding postage.

Vol. 75, No. 8 Whole No. 512, 1961

Psychological Monographs: General and Applied

THE DISTRIBUTION OF SUSCEPTIBILITY TO HYPNOSIS IN A STUDENT POPULATION:

A STUDY USING THE STANFORD HYPNOTIC SUSCEPTIBILITY SCALE[1]

ERNEST R. HILGARD, ANDRÉ M. WEITZENHOFFER, JUDAH LANDES AND ROSEMARIE K. MOORE

Stanford University

SOME people are very easily hypnotized, while others are extremely resistant. This common observation suggests that a study of the personality correlates of susceptibility might prove of interest not only in explaining the nature of the hypnotic trance, but in telling something about the organization of personality, about self-control and persuasibility.

Before attempting a study of the personality correlates we need to know whether or not we are dealing with a stable and measurable trait of hypnotic susceptibility. If on the one hand susceptibility fluctuates uncertainly from day to day, from hypnotist to hypnotist, from induction method to induction method, it would be futile to look for stable personality characteristics lying behind whimsical responsiveness. If on the other hand susceptibility turns out to be a dependable characteristic, then a search for its conditions will be promising. The investigation reported here is concerned with susceptibility as such. While personality manifestations were also studied, they are not being reported at this time. In one sense we are studying the *criterion* (hypnotic susceptibility) to be used in the later analysis of predictions to be made from variables of nonhypnotic sort. In another sense, however, this is an *aptitude* test, predicting how a person will respond in further work with hypnosis.

Variability is found in all behavior. Hence in hoping that hypnotic susceptibility may be relatively stable, we are not expecting it to be rigidly unchangeable. If there is a core of stability, we would still hope to show some circumstances under which susceptibility will change, for example, with practice, with changes in motivation, with *rapport* better established between hypnotist and subject, possibly under the influence of drugs.

What we shall mean by hypnotic susceptibility for the purposes of our investigations is a relatively persistent tendency to yield the phenomena historically recognized as belonging to the hypnotic trance, *when the opportunity to yield these phenomena is given under standard conditions.* We emphasize the importance of standard procedures at this stage, for without them we cannot hope to demonstrate such lawfulness as may exist. These conditions will not be optimal for all subjects, but later investigations can tell whether or not the conditions we have chosen are representative ones. By using such a procedure we can assign some sort of numerical value that will state the degree of susceptibility of one person relative to another under these standard conditions. By repeating the procedures on more than one day, with alternate forms, the stability of the measurement can be determined in the form of a reliability measure.

The importance of such measures for carrying on the study of personality correlates is evident. It is futile to attempt predictions if there is nothing measurable and stable to predict. But there are many other purposes for such measures. Now that hypnosis is being used widely in medical practice, as in dentistry, obstetrics, surgery,

[1] This is a report of one of a series of investigations conducted within the Laboratory of Human Development under a grant from the Ford Foundation to Robert R. Sears and Ernest R. Hilgard.

psychotherapy, it is desirable to know whether or not the hypnotic method is appropriate for a given subject. We do not now know, for example, *how much* susceptibility is desirable for the hypnotic method to be efficacious for these various purposes. It is quite possible, for example, that a very slight amount of susceptibility is all that is necessary for some psychotherapeutic uses of hypnosis. Having a standardized scale will help to make our knowledge more precise. It may also turn out that with the high motivation of a person seeking relief from pain, hypnotic susceptibility will increase. If this should be the case, it will give us important information about the nature of susceptibility. All of this needed information depends upon sound measures of susceptibility.

If our scores correlate with such measures as hypnotic amnesia, hallucination, and the carrying out of posthypnotic suggestions, we shall assume that they are sampling susceptibility to hypnosis. Low scores on our scale will be made by people who give some of the responses commonly used in tests of waking suggestion, while giving few, if any, of the responses more particularly associated with the recognized hypnotic trance; high scores will be made by those who give more of the typical trance phenomena. To serve as a predictive scale our scores need only indicate that the subjects with high scores are good candidates for hypnosis; for this purpose they need say nothing about how deeply the subjects have already been hypnotized.

A susceptibility scale can be validated by bringing back subjects who score high and low and then attempting to elicit from them other kinds of trance phenomena. Some experiments along these lines will be reported indicating that the scale appears to be valid as well as reliable.

Before presenting the data from our investigation we shall review some of the history of the problem of hypnotic susceptibility, along with the previous attempts to measure individual differences in susceptibility.

Prior Investigations of Susceptibility

The conception that hypnosis occurs in degrees has been forced upon all investigators, even though their definitions of hypnosis, or their preferences regarding it, may have led them to wish for all-or-nothing manifestations. Braid (1843) characterized the true state of hypnotic sleep according to complete spontaneous amnesia for all events occurring during the trance, but he was troubled by finding many of his patients helped by his procedures even though they failed to meet his criterion. In the end, he stood by his guns, saying that his patients were in other states, but not truly hypnotized unless completely amnesic. Charcot (1882) and his co-workers Richer (1885) and Gilles de la Tourette (1889) specified three kinds of hypnotic state (catalepsy, lethargy, and somnambulism). These were thought of as discrete, with sharp transitions between them. It was an easy further step, however, for writers such as Pitres (1891), influenced by Charcot, to add other borderline, mixed, and incomplete states (*états frustes*). The implication is still that of a mixture of states, rather than a true continuum, but once there are enough borderline conditions there is little distinction between a mixture and a continuum.

Nineteenth Century Depth Scales

The analogy with sleep makes the notion of degrees of hypnosis, expressed as degrees of depth, a plausible one. This manner of thinking seems to have been first proposed by Richet (1884). He recognized that the induced somnambulism of the mesmerists was the same as that produced by other methods, and he rejected the animal magnetism explanation. Three degrees that he recognized were: (*a*) *torpor*, in which the eyes close spontaneously and can be opened with great difficulty, if at all; (*b*) *excitation*, with total inability to open the eyes, unresponsiveness except to the hypnotist, some "automatism" and "double-consciousness"; (*c*) *stupor*, with previous phenomena in greater degree, spontaneity totally lacking, subject a complete automaton,

easily produced "contractures" and "catalepsies." There is usually amnesia in the second stage, more complete amnesia in the third. Here we have the beginning of a depth scale.

Not long afterwards, Liébeault (1889) proposed a six-point scale, and Bernheim (1891) a nine-point scale. These are summarized in Tables 1 and 2. Liébeault felt that his scale was unidimensional in the sense used much later in Guttman-type scales, that is, that an individual who showed the symptom characteristic of one of his degrees of depth would always show all the symptoms of lesser degree. Both scales emphasize spontaneous amnesia as a characteristic of the deeper stages. This is equally true for Bernheim, despite his theoretical position that all phenomena of hypnosis are the result of suggestion. Perhaps he interpreted amnesia as a natural accompaniment of other suggestions, although it was not itself suggested.

TABLE 1

DEPTH OF HYPNOSIS ACCORDING TO LIÉBEAULT (1889)

Light sleep		
1.	Drowsiness.	Torpor, drowsiness, heaviness, of the head, difficulty in opening the eyes.
2.	Light sleep.	Above signs plus catalepsy, but with ability to modify the position of members if challenged.
3.	Light sleep: deeper.	Numbness, catalepsy, automatism. The subject is no longer able to interfere with rotary automatism.[a]
4.	Light sleep: intermediate.	In addition to catalepsy and rotary automatism, the subject can no longer attend to anything else but the hypnotist and has memory only for the interchange between them.
Deep or somnambulistic sleep		
5.	Ordinary somnambulistic sleep.	Total amnesia on waking. Can have hallucinations during sleep. Hallucinations vanish with waking. Subject submits to the will of the hypnotist.
6.	Profound somnambulistic sleep.	Total amnesia on waking. Hypnotic and posthypnotic hallucinations possible. Complete submission to the hypnotist.

[a] Catalepsy refers to waxy flexibility, in which the arms remain where they are placed. Rotary automatism refers to the persistence of rotary movement of the hand and forearm, once set into motion by the hypnotist.

TABLE 2

DEPTH OF HYPNOSIS ACCORDING TO BERNHEIM (1891)

Memory retained on waking	
Degree 1.	Torpor, drowsiness, or various suggested sensations such as warmth, numbness.
Degree 2.	Inability to open the eyes if challenged to do so.
Degree 3.	Catalepsy suggested by the hypnotist and bound up with the passive condition of the subject, but may be counteracted by the subject.
Degree 4.	Catalepsy and rotary automatism which cannot be counteracted by the subject.
Degree 5.	Involuntary contractures and analgesia as suggested by the hypnotist.
Degree 6.	Automatic obedience; subject behaves like an automaton.
Amnesia on waking	
Degree 7.	Amnesia on waking. No hallucinations.
Degree 8.	Able to experience hallucinations during sleep.
Degree 9.	Able to experience hallucinations during sleep and posthypnotically.

With the appearance of such scales it became meaningful to speak of the distribution of susceptibility according to the depth of hypnosis that could be reached. It is possible to make a distinction between susceptibility and depth, but as a beginning it was natural to define susceptibility by the greatest depth that the subject was able to attain.

Many of those who worked with hypnosis were satisfied with simpler classifications of hypnotic states. A common three-point scale distinguished between "somnolence," "light sleep," or "hypotaxy," and "deep sleep" or "somnambulism." This classification was used by Forel, Loewenfeld, Fontan, Ségard, and Ringier. Others preferred a twofold classification, whereby individuals fell in

Group I if only their motor behavior was affected, and into Group II if in addition they yielded also perceptual and ideational changes. Gurney, Delboeuf, Hirschlaft, and Dessoir preferred this scheme. Mayo (1852) made a distinction between "waking" and "sleep" behavior.

These nineteenth century scales have enough in common that it is possible to make some comparisons among the findings of the various authorities. All, for example, give a good deal of weight to spontaneous ("nonsuggested") posthypnotic amnesia as a criterion of deep hypnosis. Other stages are usually described according to classes of events, rather than according to specific tests, so that there is an element of uncertainty about borderline states. Induction procedures were not standardized, except within master-disciple groups, and there was always a certain amount of accepted folklore. For example, it was assumed by many hypnotists that hallucinations were produced by a simple posturing of the subject, without verbal suggestions of hallucination. The word "suggestion" to one hypnotist might mean a verbal command, while to another it might mean a nonverbal suggestion produced by some sort of manipulation. Thus in comparing the distributions of susceptibility as reported by these early writers, one naturally must recognize large elements of uncertainty in making quantitative comparisons.

It comes as something of a surprise to find the very large numbers of subjects for whom records were kept and reported in the latter part of the nineteenth century. In Table 3 we have digested the results from two major reviews (Loewenfeld, 1901; Schmidkunz, 1894) adding some cases reported a little later by Bramwell (in 1903). The 14 summarized distributions in this table are based on records from 19,534 patients—a very substantial number, even with allowance for some duplications in the reports. There are included only those reports which permitted classification (always with a margin of uncertainty) into refractory or nonsusceptible subjects and three degrees of susceptibility: drowsy-light, hypotaxy-moderate, and somnambulism-deep. Because the conditions of each investigation differ, we have reported the

TABLE 3

DISTRIBUTION OF SUSCEPTIBILITY TO HYPNOSIS: NINETEENTH CENTURY STUDIES

Investigator	Source	Date	Age range	Sessions (N)	Cases (N)	Distribution of susceptibility (in percent)				
						Refractory: nonsusceptible	Drowsy-light	Hypotaxy-moderate	Somnambulistic-deep	Total susceptible
Peronnet	a	ante–1900			467	25	10	20	45	75
Forel	a	ante–1898			275	17	23	37	23	83
Lloyd-Tuckey	a	ante–1900			220	14	49	28	9	86
Bramwell	b	ante–1900	4–76	$M = 23$	200	11	24	26	39	89
Von Schrenck-Notzing	a	ante–1900			240	12	17	42	29	88
Mosing	c	1889–93	a few children	$M = 20$–30	594	12	42	17	29	88
Hilger	a	ante–1900			351	6	20	42	32	94
Von Schrenck-Notzing (pooling of 15 reports)	a	1892			8,705	6	29	50	15	94
Liébeault	a	1884–89	7–63		2,654	5	22	62	11	95
V. Eeden & v. Renterghem	a	1887–93			1,089	5	43	41	11	95
v. Renterghem	a	ante–1900			414	4	52	33	11	96
Wetterstrand	a	1890		Failures, 1 or 2 trials	3,209	3	36	48	13	97
Velander	a	ante–1900			1,000	2	32	54	12	98
Vogt	a	ante–1900			116	0	2	13	85	100
Total cases					19,534					
Range of percentages						0–25	2–52	13–62	9–85	75–100
Mean of percentages						9	29	36	26	91

[a] Loewenfeld (1901).
[b] Bramwell (1903).
[c] Schmidkunz (1894).

means of the investigations without respect to the variation in numbers of cases, thus using each report as one case in computing the means at the bottom of Table 3.

More Recent Scales of Hypnotic Susceptibility

A new interest in hypnotic susceptibility came to the fore around 1930, when M. M. White (1930) published his scale. He made use of specific responses to suggestions given in hypnosis as a means of arriving at scores, and thus began a practice adopted by most of the later scales. Shortly thereafter the Davis and Husband (1931) scale appeared, which, while more detailed and covering a wider range of depth, assigns scores on the basis of responses to classes of suggestions rather than to specific responses. At about the same time Barry, MacKinnon, and Murray (1931) proposed a scale based on a short list of specific suggestions. They placed much weight upon the subject's ability to have some suggested posthypnotic amnesia, and upon suggested inhibition of response, that is, loss of ability to control certain types of movement, such as separating interlocked fingers. While Hull (1933) did not develop a scale of susceptibility, he often used speed of eye closure to suggestion as a measure of susceptibility. The well-known scale of Friedlander and Sarbin (1938) combines this emphasis upon eye closure with the kinds of items used in the Barry, MacKinnon, and Murray scale. The scale developed by Eysenck and Furneaux (1945) is similar in many respects to that of Friedlander and Sarbin, while the scales of LeCron and Bordeaux (1947) and of Watkins (1949) are more nearly variations of the Davis-Husband type of scale. Of these, the scale by LeCron and

TABLE 4

DISTRIBUTION OF SUSCEPTIBILITY TO HYPNOSIS: MORE RECENT STUDIES

Investigator	Date	Subjects	Sessions (*N*)	Cases (*N*)	Distribution of susceptibility (in percent)				
					Refractory nonsusceptible	Drowsy-light	Hypotaxy-moderate	Somnambulism-deep	Total susceptible
Eysenck & Furneaux	1945	Neurotic patients	1	60	37[a]	38	17	8	63
Friedlander & Sarbin	1938	College students	1	57	33[b]	25	37	5	67
Weitzenhoffer	1956	College students	1	200	23[b]	59	15	3	77
Barry, MacKinnon, & Murray	1931	College students	1	73	16[c]	37	29	18	84
Hilgard, Weitzenhoffer, & Gough	1958	College students	1	74	3[b]	51	30	16	97
Total cases				464					
Range of percentages					3–37	25–51	15–37	3–16	63–97
Mean of percentages					22	42	26	10	78

[a] Nonsusceptible include lowest category reported, scores 0–10 on an 80-point scale; others categorized according to thirds of remaining scale (i.e., 10–80).

[b] Converted according to Friedlander-Sarbin practice, with nonsusceptible scoring 0 on Friedlander-Sarbin scale, light = 1–9, medium = 10–14, deep = 15–20.

[c] Converted according to most comparable categories.

Bordeaux is of great length, and covers a very large variety of hypnotic phenomena.

The most widely used of these scales have been those of Davis and Husband and of Friedlander and Sarbin. Because descriptions of these scales are readily available elsewhere (e.g., Weitzenhoffer, 1957) they will not be repeated here.

Results of investigations using the more modern scales are presented in Table 4. For this purpose, responses have been reclassified in the categories of Table 3, with every effort to be as fair as possible to the new and old conceptions. Comparison of Tables 3 and 4 shows that the ranges within each category of depth overlap substantially. Recent investigators report a higher mean percentage of refractory subjects and a lower mean percentage of somnambulistic subjects, but the orders of magnitude are similar. The older investigators were concerned primarily with clinic patients, for whom the motivation for successful hypnosis was very high, while the modern investigators (with the exception of Eysenck and Furneaux) used college students. There were some children, also, in the older samples, and as we shall see, they are more susceptible than adults. Also, of course, the older studies often used many sessions. In any case results are sufficiently variable, even among the three studies using a modern scale such as the Friedlander-Sarbin one, that further studies are clearly in order.

Form of Distribution

The average results, when classified according to the nineteenth century categories, suggest a fairly normal distribution of hypnotic susceptibility, with a few refractory cases, a few very good subjects, and the rest falling between. When, however, results are scaled according to standardized measurements, various forms of distribution are reported. Perhaps the most commonly reported distribution is that of an inverted J, with most subjects relatively little susceptible, and a pronounced skew of the distribution toward the more susceptible end of the scale (e.g., Eysenck & Furneaux, 1945; Friedlander & Sarbin, 1938; Hilgard, Weitzenhoffer, & Gough, 1958; Weitzenhoffer, 1956). With a somewhat different scaling of scores, however, a bimodal distribution may result (Hilgard et al., 1958). There are traces of bimodality in a number of previously published investigations (e.g., Barry et al., 1931; Davis & Husband, 1931). The uncertainty about the form of distribution raises a number of questions, to which we shall return.

Age and Susceptibility

There is a good deal of incidental evidence that children make unusually good hypnotic subjects (e.g., Bramwell, 1956) but careful investigations are lacking. The one best investigation was carried out by Liébeault (as reported by Beaunis, 1887). He studied 744 subjects ranging from early childhood to old age, with the results shown in Figure 1. Of his child subjects below the

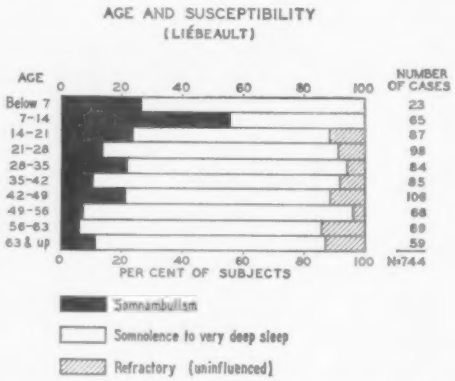

Fig. 1. Age and susceptibility to hypnosis. Data from Liébeault as reported by Beaunis, 1887.

age of 14, none was uninfluenced, compared with about 10% uninfluenced at most other ages. The highest proportion of somnambulistic subjects was found between the ages of 7 and 14 (55.3%). There is little progressive change in susceptibility beyond the age of 14. The conclusions of Ringier (Schmidkunz, 1892) and of Mosing (Schmidkunz, 1894) support Liébeault's findings.

While comparable data are not available from more recent studies, one of Hull's stu-

dents, Ramona Messerschmidt, reported data on postural sway for a sample of children between the ages of 5 and 16. Because postural sway shows a positive correlation with hypnotizability, these data are relevant. She found an increase in responsiveness from age 5 to high points at ages 6 and 8, with slow decline thereafter (Hull, 1933, p. 84). Her results are also coherent with those of Liébeault.

Sex and Susceptibility

Perhaps because of the early association of hypnosis with hysteria, and of hysteria with the female, there is a popular misconception that women are more readily hypnotized than men. The problem has been of continuing interest, and there have been numerous reports concerning sex differences. Wetterstrand, (Loewenfeld, 1901) felt there were no sex differences, a conclusion with which Bramwell (1956) agreed. Liébeault reported his findings carefully, finding no sex differences. The occasional reports of differences favoring women have found differences too slight to meet tests of statistical significance (Davis & Husband, 1931; Friedlander & Sarbin, 1938; Weitzenhoffer & Weitzenhoffer, 1958). One exception is the report of Hilgard et al. (1958) reporting a statistically significant difference, with women more susceptible.

Not only are mean sex differences seldom found, but the distributions are very similar for men and women, even when circumstances of experimentation and measurement produce distributions differing in general form from one study to another.

Summary

Differences in susceptibility to hypnosis have been found by all workers. Although the detailed findings are in disagreement, the nineteenth century studies, carried on with thousands of patients, are in general agreement with the more recent laboratory studies done largely with college students. The earlier studies tend to report somewhat higher average success, but many factors can account for this: the motivation of the subjects, the inclusion of children, repeated hypnotic sessions. The basic phenomena considered to be signs of the hypnotic trance are very much today what they were then. A particular thread of continuity is the use of posthypnotic amnesia as a sign of a substantial trance.

The early study of Liébeault on age difference remains essentially sound, although further studies should be conducted. It appears from what evidence is available that hypnotic susceptibility is at its height somewhere between the ages of 7 and 14.

Sex differences are not at all prominent, if indeed they can be demonstrated at all. There may be a slight tendency for women to be more susceptible than men, but if the tendency exists it is slight indeed.

There are other problems of hypnotic susceptibility, such as the dimensions of hypnotizability. These we shall postpone for later consideration.

Stanford Hypnotic Susceptibility Scale

A long-range study of hypnotic susceptibility was begun at Stanford during the academic year 1957–58. A preliminary study of individual differences in susceptibility, based on the scores of 74 subjects (Hilgard et al., 1958) pointed up the need for some revisions of the scale then used, a slight modification of the Friedlander-Sarbin scale (1938). A revision of the scoring weights led in that study to a marked bimodality of scores, a result that it was feared might have been due to the nature of the items in the scale. Also the very low scores (of 0 and 1) were so frequent that it seemed desirable to add some easier items in order to spread out the score distribution.

With this background a new susceptibility scale was prepared and after some pretesting in the summer of 1958 a standardization test was begun in the autumn of 1958. This report is based on the results of 124 subjects tested with the new scale. Because additional scales are in preparation, the initial scale is known as the Stanford Hyp-

TABLE 5

ITEMS IN THE STANFORD HYPNOTIC SUSCEPTIBILITY SCALE

Item	Form A	Form B	Criterion of passing
Postural sway	Backwards	Backwards	Falls without forcing
Eye closure	Form A Induction	Form B Induction	Closes eyes without forcing
Hand lowering	Left	Right	Lowers at least 6 inches by end of 10-second timed interval
Immobilization	Right arm	Left arm	Arm rises less than 1 inch in 10-second timed interval
Finger lock	Before chest	Overhead	Incomplete separation of fingers at end of 10 seconds
Arm rigidity	Left arm	Right arm	Less than 2 inches of arm bending in 10 seconds
Hands moving	Together	Apart	(A) Hands as close as 6 inches; (B) Hands at least 6 inches apart
Verbal inhibition	Name	Hometown	Name unspoken in 10 seconds
Hallucination	Fly	Mosquito	Any movement, grimacing, acknowledgment of effect
Eye catalepsy	Both eyes closed	Both eyes closed	Eyes remain closed at end of 10 seconds
Posthypnotic	Changes chairs	Rises, stretches	Any partial movement response at posthypnotic signal
Amnesia test	Recall of items—3 to 11	Recall of items—3 to 11	Recall of 3 or fewer items

notic Susceptibility Scale which is Part I of the total Stanford Hypnotic Scales. Part II will be a scale sampling a greater variety of hypnotic phenomena, for subjects who prove susceptible on Part I. The susceptibility scale has been separately published (Weitzenhoffer & Hilgard, 1959). The published version gives two forms of the scale, with complete directions, and some preliminary standardization data. Hence the detailed instructions will not be repeated here.

The 12 items that receive scores in the scale are summarized in Table 5. An additional item (arm catalepsy after passive lifting of the arm) was tested but dropped from the final scale because of its failure to correlate with the other items.

Equivalence of the Two Forms. For the purpose of standardization, the two forms of the scale were given, half in the order A-B and half in the order B-A. There were no significant differences between forms, or between the scores on 2 days of hypnosis. For most of the analyses of this report the two forms will be considered as merely two halves of one total test, and the orders in which the forms were given will be ignored.

Reliability. The reliabilities as determined for each day separately, for the total 2-day score, and the retest reliability, are given in Table 6. Because the reliabilities for 1-day scores are .83, the retest reliability, on the assumption of perfect correspondence between the 2 days, should also be .83, as indeed it does turn out to be. Also, if the score of each of the days is considered to be but half of a total test, the Spearman-Brown formula would predict a reliability of the total 2-day test of .91, as indeed it is found to be by the Kuder-Richardson formula (their Formula 20). Hence the interrelationships of the coefficients of Table 6 are consistent with the interpretations being made of the test results. The most satisfactory reliability is obviously that of the total 2-day scores (.91). This is high enough to permit the use of these scores as criteria for other purposes.

The changes in score between the 2 days are of some interest. Therefore a scatter-

TABLE 6

RELIABILITY COEFFICIENTS: STANFORD HYPNOTIC SUSCEPTIBILITY SCALE

Type of reliability	Cases (*N*)	Reliability coefficient
Kuder-Richardson		
Day 1 (Forms A,B)	124	.83
Day 2 (Forms A,B)	124	.83
Days 1 + 2 (Forms A + B)	124	.91
Retest		
Day 1 (Form A) vs. Day 2 (Form B)	60	.78
Day 1 (Form B) vs. Day 2 (Form A)	64	.87
Day 1 (Forms A,B) vs. Day 2 (Forms A,B)	124	.83

plot of the 2-day performances is given in Figure 2. Analysis shows that 68% of the cases have a second day's score within one point of the first day, and 93% have scores

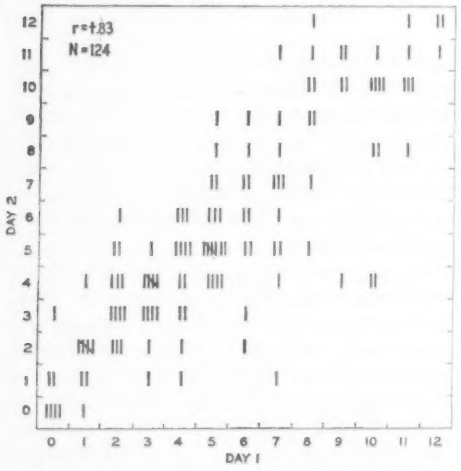

FIG. 2. Scatterplot of scores on 2 days of hypnosis. (The days followed each other at 24- or 48-hour intervals; no differences attributable to the interval were detected.)

on the 2 days within three points of each other.

Validity. The validity of the scale has been tested by bringing back to the laboratory subjects from both the high and the low end of the distribution.

The Stanford Hypnotic Susceptibility Scale is the first of several scales being developed in this laboratory. For convenience we may refer to it as Part I, and talk about the later scales as Parts II and III. The next scale beyond Part I is designed to sample a more varied set of behaviors under hypnosis, so that, in effect, it will spread out the better subjects of Part I into those who can go much further and those who have gone about as far in hypnosis as they are ready to go. Although this scale, known as Part II—Extended Susceptibility Scale, is undergoing revision, a preliminary form was being tested in 1958–59, when the present sample was collected. Its general nature can be inferred from the 15 items that entered into the scores of this preliminary form. Each of these items was scored on a pass-fail basis, just as in the scoring of Part I:

1. Inability to stand up
2. Anesthesia (hand)
3. Taste hallucination
4. Smell hallucination
5. Heat illusion
6. Music hallucination
7. Visual hallucination of sport event
8. Regression to a recently seen motion picture
9. Dreaming under hypnosis
10. Visual hallucination of record player (eyes open)
11. Suggested deafness: inability to hear tapping sounds
12. Negative visual hallucination: missing clock hand
13. Posthypnotic suggestion
14. Amnesia
15. Reinduction of hypnosis at a signal

In connection with the development of the scales of Part II, we sought to rehypnotize as many as possible of the subjects who scored high on the susceptibility scale. We succeeded in bringing back for further study 21 of 28 or 75% of the highest scoring subjects, all having scored 16 points or more of the possible 24 points on the first 2 days. Of these 21 subjects, 15 (72%) scored in the upper half of the Part II scale (Table 7). We also tested a few subjects who had not scored as high on the susceptibility scale, but were potentially promising because they had shown considerable amnesia. There were 21 of these subjects, rep-

TABLE 7

RELATIONSHIP BETWEEN SCORES ON PART I (SUSCEPTIBILITY SCALE) AND SCORES ON PART II (DEPTH SCALE)

Part I (Susceptibility) (Days 1 and 2 combined)	Part II (Depth)		
	Low scores (0–7)	High scores (0–14)	Total
High scores (16–24)	6	15	21
Medium scores (7–15)	16	5	21
Total	22	20	42

Chi square = 9.55; $.01 > p > .001$.

Note.—The Depth Scale is not yet in form for publication.

resenting a selected third of the 62 subjects scoring between 7 and 15 points out of the 24 possible on the first 2 days. Of these 22 subjects, only 5 (24%) scored in the upper half on Part II. We thus see that the better subjects on the susceptibility scale were a more promising pool for providing good subjects on Part II than the poorer subjects on the susceptibility scale. This gives us some confidence in the validity of the measures, although the occasional exceptions require further study. We shall turn later to some questions of dimensionality raised by these exceptions.

A number of subjects who scored very low on the susceptibility scale were also invited back individually for a further hynotic session.[2] In these cases the hypnotist departed entirely from standard procedures, and did his best to capitalize on the known successes of the subject on the prior days, and on any cues picked up during the attempted induction. Review of the protocols of the 17 subjects tested in this way (subjects who scored 0–11 on 2 days, with a median of 7 out of a possible 24), showed performances on this third day with no greater success than would be predicted from the first day scoring level. There were two doubtful cases, in which responsiveness appeared to increase, but neither of these cases demonstrated amnesia or responded to posthypnotic suggestion. Hence it appears that subjects refractory on the susceptibility scale, using the standard method of induction, were also refractory when another method was used. Generalization from this finding is limited because only one additional hour was used in the attempted reinduction.

These results give us some confidence that our susceptibility scale is both reliable and valid. The results have to be interpreted with some caution because they are limited to a sample of college students, no long-term hypnosis was tried, and we did not manipulate the motivation of our subjects. But within these limitations, we have a scale that is dependable enough to be used in the study of the lawfulness of hypnotic phenomena.

POPULATION STUDIED

The 124 university undergraduates (64 men and 60 women) whose scores are here considered were volunteers fulfilling part of the course requirement in introductory psychology at Stanford. We need to ask: What kinds of students come to Stanford? How representative are our subjects of the Stanford population?

As a private university with high admission standards and high tuition fees, Stanford attracts students who represent a selection largely from upper middle class homes, although the prevalence of scholarships and work opportunities means that there are many bright students from lower socioeconomic strata as well. Stanford is coeducational, with some 5,000 undergraduate students of whom about 1,700 are women.

The course in introductory psychology is a popular one, so that about two-thirds of the students take the course at some time during their undergraduate years. It is chosen by engineering and other preprofessional students as well as by the liberal arts students. Comparison of the majors of our hypnotic subjects with the distribution of majors in the university at large shows that

[2] We wish to acknowledge our indebtedness to Jay D. Haley of the Palo Alto Veterans Administration Hospital for assistance in this portion of the experiment.

TABLE 8

UNDERGRADUATE MAJORS OF HYPNOTIC SUBJECTS COMPARED WITH DISTRIBUTION OF MAJORS IN THE TOTAL UNDERGRADUATE BODY

Major	Male students who have designated majors		Female students who have designated majors	
	Total undergraduate (%)	Hypnotic sample (%)	Total undergraduate (%)	Hypnotic sample (%)
Humanities and Sciences				
Humanities	8	15	28	16
Social sciences	35	36	43	58
Physical and biological sciences	17	11	12	2
Engineering and Mineral Sciences	38	25	1	0
Professional and Preprofessional	2	13	16	24
Total	100	100	100	100
Number of cases	2,359	55	1,109	55
Major not chosen	903	9	568	5
Grand Total	3,262	64	1,677	60

we have a reasonable cross section of student interest represented (Table 8).

We may assume that from the point of view of demographic characteristics our sample is fairly representative of the Stanford undergraduate. The question remains whether or not any subtle bias enters through the process of volunteering for an experiment involving hypnosis. We have attempted to get an index to this bias by having members of the class volunteer first for a session in which they have an opportunity to complete a personality inventory, and then we have limited the group acceptable for hypnosis to those who first volunteered for the personality inventory. If there are personality factors leading some kinds of people to volunteer for hypnosis, and others to refrain from volunteering, we should be able to detect some differences in their scores on the personality inventory. There were limited opportunities to participate in the hypnotic experiments, and many of those listed as nonvolunteers would gladly have participated. Hence it must not be assumed that any large fraction of the nonvolunteers had qualms about hypnosis. Some illustrative results are given in Table 9.

The means and standard deviations come very near to the norms as published for the two scales reported (Gough, 1957). For example, for a sample of 680 college students the Dominance scale is reported to yield a mean of 28.5 with a standard deviation of 6.0, and the Self-Control scale a mean of 29.2 with a standard deviation of 7.1. For our subjects the means for Dominance and Self-Control are 29.2 ($SD = 6.1$) and 27.4 ($SD = 6.5$), respectively.

The only evidence of any bias is in the somewhat lower scores on the Self-Control scale of our hypnotic subjects as against the nonvolunteer sample. The difference approaches significance for the female subjects ($p = .05$), and lies in the same direction for the male subjects, although the difference for them is not significant. Because within this sample the Self-Control scale correlates $r = -.39$ with our 2-day hypnotic scores for women students, any bias introduced would tend to raise the level of susceptibility of our women subjects against that of the total student sample.

In summary, the sample represents a cross section of the Stanford undergraduate student body with respect to choice of major. It is possible that there is some slight bias in the volunteer sample; if there is, it overrepresents slightly the susceptible women students.

Distribution of Susceptibility

Having established that the Stanford Hypnotic Scale yields data that show both reliability and validity, and that the sample is at least a moderately satisfactory one to represent Stanford undergraduates, we are prepared now to turn to the more detailed findings with respect to individual differences in scores.

The score distributions for the 2 days separately, and for the 2 days combined, are

TABLE 9

SAMPLE VOLUNTEERING FOR HYPNOSIS COMPARED WITH SAMPLE NOT VOLUNTEERING AMONG THOSE WHO TOOK PERSONALITY TEST[a]

Sample	Male subjects					Female subjects				
	Dominance (Do)			Self-Control (Sc)		Dominance (Do)			Self-Control (Sc)	
	N	*M*	*σ*	*M*	*σ*	*N*	*M*	*σ*	*M*	*σ*
Volunteered for hypnosis	62	29.8	6.3	27.1	6.7	59	28.7	6.0	27.7	6.3
Did not volunteer for hypnosis	92	29.2	5.3	28.2	7.2	74	28.5	5.5	30.1	7.3
Differences between means (Hyp. − nonhyp.)		0.6			−1.1		0.2			−2.4
Critical ratio		0.6			1.0		0.2			2.0*

[a] California Psychological Inventory. Palo Alto, Calif.: Consulting Psychologists Press.
* $p < .05$.

TABLE 10

RAW SCORE DISTRIBUTIONS: ALL SUBJECTS, BOTH DAYS SINGLY AND COMBINED

	Scores on single days					Scores on 2 days combined		
	Score	Day 1		Day 2		Score	Day 1 + Day 2	
More susceptible	12	3		4		24	2	
	11	6		7		22–23	3	
	10	9	31	11	33	20–21	11	28
	9	5		5		18–19	9	
	8	8		6		16–17	3	
Less susceptible	7	11		8		14–15	10	
	6	10		10		12–13	12	
	5	18		19		10–11	16	
	4	13	93	19	91	8–9	15	96
	3	12		12		6–7	17	
	2	13		11		4–5	11	
	1	9		7		2–3	8	
	0	7		5		0–1	7	
N		124		124			124	
M		5.25		5.48			10.73	
σ		3.22		3.20			6.14	

Day 1 vs. Day 2: Mean difference = .23
σ_{DM} = .17
CR = 1.35 (p = .09, one-tailed test)

presented in Table 10, along with some statistical measures based on the distributions. The change in scores between the first and second day is in the expected direction of a slight mean increase but the difference is not statistically significant.

Because there were no sex differences ($CR = 0.35$) the sexes are not treated separately. This finding disagrees with that of our first year of experimentation (Hilgard et al., 1958), when it was found that women were more susceptible than men. While a somewhat different procedure was used, we have been unable to determine any basis for the disagreement.

When the 2-day scores are plotted as a frequency distribution (Figure 3) they yield

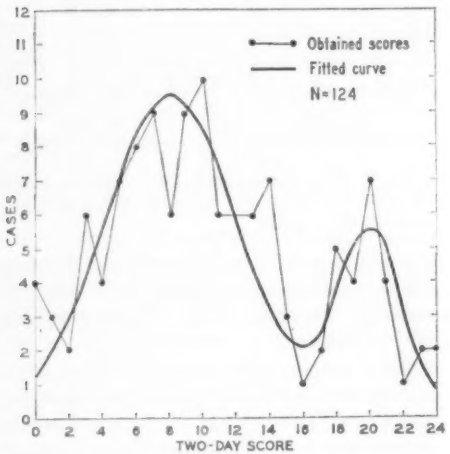

FIG. 3. Two normal curves fitted to obtained data.

a strikingly bimodal distribution. As an illustrative curve of best fit, the combination of two normal distributions has been superimposed on the obtained values in Figure 3. These were obtained by computing means and standard deviations of the scores above and below the low point as though they were separate distributions. Half the scores at the low point were assigned to the lower distribution, half to the upper. Then points on the normal curve were obtained, corresponding to the scale values of the plotted scores, and the two distributions added where they overlapped. The curve about the first mode is based on a mean of 8.07, $SD = 4.01$, $N = 96.5$, while the second mode is represented by a mean of 20.00, $SD = 1.93$, $N = 27.5$. Departure of the true points from the fitted curve is well within the limits of chance, as determined by a chi square test for goodness of fit ($p = .95$). That is, if the "true" population of scores is distributed according to the fitted curve, deviations of the size found would be expected 95 out of 100 times for a sample of this size.

A single normal curve can also be fitted to the data without doing much violence, because the second mode is a small one. The departure of the obtained scores from a best fit normal curve is not bad by the chi square criterion ($p = .50$), so that it is permissible to enter the scores into correlations.

The bimodality of scores is consistent with the findings of the earlier investigation (Hilgard et al., 1958). Because bimodality is rather unusual in psychological investigations, the bimodality of the distributions deserves special consideration.

The bimodality of score distributions, found in two successive yearly studies using somewhat different procedures, raises interesting questions. Does it mean that there is some sort of "type" distribution underlying susceptibility to hypnosis, or does it mean that some procedure in assembling test items and in weighting scores has produced the bimodality?

Problem of Bimodality

We may begin with the assumption that the distribution of susceptibility is not inherently bimodal, and show, under this assumption, how bimodality can arise. That bimodality is not inevitable has already been made clear in our earlier study in which, using the Friedlander-Sarbin weights the distribution was in the form of an inverted J, while by using dichotomized scores the distribution became bimodal (Hilgard et al., 1958). This finding has led us to examine some of the things that happen when scores are dichotomized. The following principles are known to statisticians, but they are seldom of interest to psychologists because

they apply primarily to scales built of items that correlate higher than most psychological scores correlate.

1. *If dichotomized items of equal difficulty are combined into a scale by simple addition, the higher the intercorrelations between the items, the greater the bimodality of the resulting distribution.* It is possible to begin with items that split exactly 50–50, so that there is no inherent tendency to prejudice the form of distribution, and then to combine them to form various kinds of symmetrical distribution. The higher their intercorrelations, the more evident the bimodality of their composite scores will be (e.g., Guilford, 1950, p. 491).

2. *If highly intercorrelated dichotomized items are of unequal difficulty, the form of the distribution of scores based on adding the item scores will depend upon the distribution of the item difficulty.* The simplest way to demonstrate this empirically is to arrange some artificial scores into Guttman-type scales, constructing the scale to yield the highest possible item intercorrelation for dichotomized items varying in difficulty. The bimodality that would result if the items were of equal difficulty need not result if the items are of sufficiently different difficulties. Some illustrations are given in Figure 4. It is easy to see how manipulation of item difficulty can produce any desired form of distribution.

The demonstrations of Figure 4 show that with high correlations and some choice in item difficulty, almost anything can happen to the form of score distribution. Therefore one must be very careful not to assert anything about the distribution of the phenomena underlying the scores unless more information is available than a distribution of scores added up from dichotomized items.[3]

Both the principles above apply to our data, for we have a number of highly intercorrelated items, and they vary in difficulty.

DIFFICULTY DISTRIBUTION I

Subject	Item Scores a	b	c	d	Total Score
A	+	+	+	+	4
B	+	+	+	+	4
C	+	+	+	+	4
D	+	+	+	−	3
E	+	+	+	−	3
F	+	+	−	−	2
G	+	−	−	−	1
H	+	−	−	−	1
I	−	−	−	−	0
J	−	−	−	−	0
K	−	−	−	−	0
L	−	−	−	−	0
% Success	67	50	42	25	46

DIFFICULTY DISTRIBUTION II

Subject	Item Scores a	b	c	d	Total Score
M	+	+	+	+	4
N	+	+	+	−	3
O	+	+	+	−	3
P	+	+	+	−	3
Q	+	+	−	−	2
R	+	+	−	−	2
S	+	+	−	−	2
T	+	+	−	−	2
U	+	−	−	−	1
V	+	−	−	−	1
W	+	−	−	−	1
X	−	−	−	−	0
% Success	92	67	33	8	50

FIG. 4. How distribution of item difficulty affects the form of distribution of scores based on highly intercorrelated items.

By selecting items of appropriate difficulty (or by dichotomizing them somewhat differently) we could obtain almost any form of distribution.

True bimodality may exist. Despite the caution that is needed in asserting anything about the "true" form of distribution, one item in our scale is measured by a "natural" kind of scale, so that the form of distribution can be determined for this item with a minimum of artifacts due to scaling. This is the item which tests recall under suggested amnesia. The form of the interrogatory at the end of the hypnotic session permitted the subject to recall from 0 to 10 items, and his raw score is simply the number of items recalled. There are no constraints on this score that should make it bimodal, unless there is something about the distribution of amnesia that is inherently bimodal. The distribution is plotted in Figure 5. The scores have been converted from items recalled to items forgotten, in order to make the form of distribution conform to that used in the distribution of suscepti-

[3] Eysenck and Furneaux (1945) in discussing the problem of score distribution consider the consequences of cutting points on otherwise normally distributed scores. Whenever scores are dichotomized, some statistical problems arise.

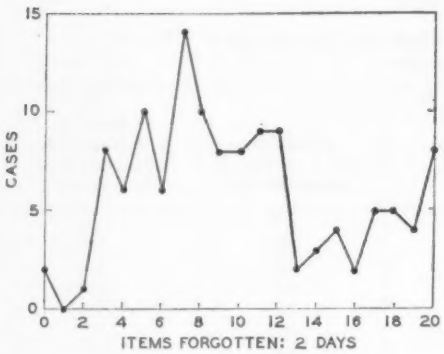

FIG. 5. Distribution of amnesia scores obtained following each of 2 days of hypnotic induction. ($N = 124$)

bility in Figure 3. The distribution is strikingly bimodal, and the number of cases in the upper mode (33 of 124) is not far from the 28 of 124 in the upper mode of the total susceptibility scale (Table 10). These results keep alive the possibility that the true distribution of hypnotic susceptibility in our population is bimodal.

The only way in which the choice of items could bias the form of the amnesia distribution would be by way of some causal influence upon recall of the items as a consequence of experience with them. The correlation between other scale items and amnesia might conceivably result in a similarity in form between the distribution of susceptibility and the distribution of amnesia. In that case the resultant susceptibility is in reality bimodally distributed, but its distribution might be artificially produced by the kinds of items used in the test. Actually we shall elsewhere report some analyses showing that recall is in fact affected by the nature of the items, but the tendency is to have somewhat less amnesia for successful than for unsuccessful items, when depth of hypnosis is taken into account (Hilgard & Hommel, 1961). Any similarity between the forms of distribution for susceptibility and for amnesia must therefore depend upon fairly complex underlying processes.

We leave the problem of bimodality here. On the one hand we have noted conditions under which the form of the curve of distribution can be artificially manipulated, which leads us to distrust the bimodality of our scores derived from summing dichotomized items. On the other hand, the presence of bimodality in the amnesia scores suggests the possibility of some genuine bimodality underlying the results.

Problem of Dimensionality

There have often been suggestions that there are different kinds of hypnotic susceptibility, as reflected in the differential passing of unlike items. Thus White (1937) proposes that there are two kinds of hypnotic trance (active and passive) with different personality correlates. Eysenck and Furneaux (1945) propose a distinction between primary and secondary suggestibility, as measured by different items; they also suggest a difference between active and passive subjects. It is important to examine our criteria measures, therefore, to determine if there is any evidence for more than one kind of ability underlying hypnotic susceptibility.

It should be pointed out that hypnosis itself is classified by Eysenck and Furneaux as illustrating "primary" suggestibility, and the tests that correlate positively with it are all tests of primary suggestibility; secondary suggestibility, which they sometimes refer to as "gullibility," does not correlate with hypnosis.[4] If we should fail to find more than one kind of trait making up suggestibility to hypnosis (among items predictive of hypnosis) we would not be contradicting their findings.

Item Correlation with Total Scale. One line of evidence suggesting that there is a common dimension running through all of the items of the scale is furnished by the relatively high correlation between the score on each item and the total score minus that item. Because each item is dichotomized, and the total score can be expressed along a scale, it is appropriate to use biserial correlations (Table 11). For the scores of

[4] A study recently completed in our laboratory (Moore, 1961) lends support to this distinction. See also Stukát (1958).

TABLE 11

CONTRIBUTION OF EACH ITEM WITHIN THE TOTAL SCALE
(N = 124 throughout)

Item	Percentage passing	Reliability (Day 1 vs. Day 2) tetrachoric r's	Correlation with total scale minus this item biserial r's
Postural sway	69	.96	.38
Eye closure	58	.78	.57
Hand lowering	81	.83	.63
Arm immobilization	14	.74	.75
Finger lock	32	.83	.72
Arm rigidity	32	.88	.83
Moving hands	70	.75	.51
Verbal inhibition	23	.94	.79
Hallucination	35	.71	.55
Eye catalepsy	30	.94	.79
Posthypnotic suggestion	49	.60	.60
Amnesia	32	.77	.69

Day 1, as shown in the table, these correlations range from a low of .38 for postural sway, measured prior to induction, to a high of .83 for arm rigidity. These correlations indicate a high common factor running through the scale.

The one rejected item (arm catalepsy) was discarded because it did not correlate with the rest. The item consisted of the passive raising of the forearm with the elbow resting on the arm of the chair. It was scored as a pass if the arm remained in position rather than returning to the arm of the chair. While the 2-day retest reliability was satisfactory (r_t = .92), the biserial correlation on Day 1 with total score minus that item was −.14. Thus in part through item selection, in part through the nature of the phenomena themselves, our scale is highly saturated with what may be designated "primary suggestibility."

A Guttman-Type Scale. Guttman (1950) has proposed an arrangement of items in attitude scales to determine whether or not there is a single dimension running through the items. His type of scale can be applied to other kinds of tests provided there are high intercorrelations among the items. If items yield dichotomized scores (as ours do), then all that is necessary is to arrange the items in order of descending difficulty (i.e., percentage passing) along one axis, and subjects in descending order of total scores along the other axis. Then a plot of success by item by subject will yield a triangular distribution, with the "plus" signs concentrating in the upper left of the diagram, and the minus signs in the lower right. A fraction of such a diagram, using actual data from the first 32 subjects of our study, is presented in Table 12. Ideally, no subject should have had any successes (+'s) to the right of the solid line.

When items are arranged as in Table 12 it is possible to compute the coefficient of reproducibility as a percentage of the items falling where they should fall if the scales were perfect, that is, if there were no blanks above the stepwise line in the table. This specimen table has a coefficient of reproducibility of .92, which satisfies the Guttman criterion of .90 for a satisfactory scale. Applying this method to all our data for 124 subjects for the 24 items of 2 days of hypnosis we find a coefficient of reproducibility of .88, which, while slightly below the value designated as desirable by Guttman, comes close to it, and strongly supports the interpretation that the tests measure a unidimensional trait. The Guttman scaling is

TABLE 12

ILLUSTRATIONS OF GUTTMAN-TYPE SCALING FOR ITEMS FROM HYPNOTIC SUSCEPTIBILITY SCALE

Subject	Items from Hypnotic Susceptibility Scale (in descending order of difficulty)												Total score
	3	1	2	7	9	5	11	10	6	12	8	4	
12	+	+	+	+	+	+	+	+	+	+	+	+	12
23	+	+	+	+	+	+	+	+	+	+	+	+	12
11	+	+	+	+	+	+	+	+	+	+	+		11
24	+	+	+	+	+	+	+	+	+			+	10
5	+	+	+		+	+	+	+	+	+	+		10
3	+	+	+	+	+		+	+		+	+		9
16	+	+	+	+	+		+		+	+		+	9
8	+	+	+	+	+		+	+		+			8
20	+	+		+		+	+	+	+		+		8
7		+	+		+	+	+		+		+		7
14	+	+	+	+	+	+				+			7
15	+	+	+	+	+					+			6
6	+	+		+	+		+						5
31	+	+	+	+				+					5
2	+	+	+	+									4
19	+	+	+	+									4
22	+	+		+	+								4
27	+			+		+			+				4
29	+	+			+		+						4
32	+	+	+			+							4
10	+	+		+									3
28	+		+	+									3
4	+		+										2
18		+				+							2
21	+	+											2
25						+		+					2
13			+										1
17	+												1
26	+												1
30		+											1
1													0
9													0
Total successes per item	25	23	18	18	14	12	12	10	9	9	7	4	

Note.—Item identification (in order of presentation):

1. Postural sway
2. Eye closure
3. Hand lowering
4. Immobilization (Arm)
5. Finger lock
6. Arm rigidity
7. Hand movement
8. Verbal inhibition
9. Hallucination
10. Eye catalepsy
11. Posthypnotic suggestion
12. Amnesia

imperfect, however, and it is possible that some second dimension is interfering with unidimensionality.

Item Intercorrelations and Factor Analysis. Dichotomized items have to be intercorrelated from fourfold tables. When we plotted our data we noted that there were often empty cells, owing to the high intercorrelations and the extreme splits for some of the items. Because such tables yield indeterminate results, we decided to use 2-day scores, which fall on a three-point scale (i.e., −−, −+, ++). We then dichotomized anew in such a manner as to avoid

TABLE 13

DISTRIBUTION OF ITEM DIFFICULTY

Item	Percentage of subjects passing item		
	Day 1	Day 2	Days 1 and 2 combined[a]
Postural sway	69	81	67
Eye closure	58	67	52
Arm lowering	81	86	78
Arm immobilization	14	16	22
Finger lock	32	31	40
Arm rigidity	32	26	36
Moving hands	70	77	64
Verbal inhibition	23	19	32
Hallucination	35	34	46
Eye catalepsy	30	31	35
Posthypnotic suggestion	49	52	36
Amnesia	32	27	39

[a] Items dichotomized after combining 2 days, in order to bring cutting points nearer to 50–50.

empty cells. The resulting distribution of item difficulty is given in Table 13, in which the original item difficulties are also shown. Because of the high intercorrelations between the days it was not possible to make very drastic shifts in the cutting point (e.g., by accepting a single day's passing as enough for the more difficult items, and requiring passing on both days for the easier ones). The resulting fourfold tables of intercorrelations turned out to be without empty cells, however, so that the unambiguous use of tetrachoric correlations became possible.

A table of intercorrelations was prepared, and the table factor analyzed by Thurstone's centroid method. The intercorrelations, and the loadings on the first factor (unrotated) are given in Table 14. In order to make it easier to read the table, the items have been rearranged in the order of their loadings on this first factor. The diagonals represent the reliabilities, as determined by retest tetrachoric correlations, stepped up by the Spearman-Brown formula.

The items in Table 14 have also been grouped as "challenge" and "other" items. It is of some interest that the five items of highest loadings, all intercorrelating from .65 to .87, are of the so-called "challenge" type, in which the subject is unable to counteract the suggestions of the hypnotist, even when told to try. They are also the most difficult items, as shown in the earlier Table 13, with only two other items (amnesia and posthypnotic suggestion) falling in their range of difficulty. In some sense amnesia belongs with them, because the subject is asked to recall after being told that he will be unable to do so; this is a kind of challenge. It is not surprising, therefore, that amnesia has the next position in the table of intercorrelations.

The loadings on the first factor are high enough to account for 51% of the variance. The second and third factors account for 12% and 6% of the variance, respectively, so that the three factors account for 69% of the variance. Further factor analyses are in process, in which items better representative of the subordinate factors are included. It is evident, however, that the test as scored reflects heavily the common factor.

Thus the biserial correlations between single items and the total scale minus that item, the Guttman scaling, and the factor analysis, all point to the same conclusion that there is essentially a single dimension running through these scores, that there is basically one kind of hypnotic susceptibility being measured, although the slight differences in items undoubtedly contribute something—if nothing else, a practical method of deriving scores covering the whole range of susceptibility. The results have led us to retain the uniform weighting of all items entering into our scale. Because they have not been adjusted for variance and covariance, they are not equally weighted, but further adjustment appeared to be an unnecessary refinement.

How do these results square with what others have found? The conclusion of Eysenck and Furneaux (1945) that there were two kinds of suggestibility was cited earlier. It was pointed out, however, that their "secondary" suggestibility did not correlate with hypnosis. Because we eliminated any items not correlating with hypnosis, it is not

TABLE 14

ITEM INTERCORRELATIONS: DAY 1 AND DAY 2 COMBINED
(N = 124)

Item	I. Challenge items					II. Other items							Loading on first factor
	6	5	4	10	8	12	2	3	9	7	11	1	
I. Challenge items													
6. Arm rigidity	[.94]	.87	.82	.80	.75	.63	.54	.59	.48	.29	.48	.22	.86
5. Finger lock		[.92]	.65	.78	.90	.73	.58	.44	.48	.30	.22	.30	.85
4. Arm immobility			[.84]	.78	.70	.60	.55	.38	.49	.18	.61	.10	.78
10. Eye catalepsy				[.97]	.76	.32	.51	.46	.46	.40	.22	.35	.78
8. Verbal inhibition					[.97]	.58	.42	.42	.39	.29	.21	.23	.77
II. Other items													
12. Amnesia						[.84]	.50	.51	.66	.29	.32	.42	.74
2. Eye closure							[.90]	.46	.55	.50	.35	.42	.70
3. Hand lowering								[.89]	.36	.71	.48	.40	.69
9. Hallucination									[.82]	.50	.54	.37	.69
7. Moving hands										[.84]	.58	.55	.62
11. Posthypnotic											[.77]	.30	.59
1. Postural sway												[.96]	.49

surprising that we did not find secondary suggestibility. Our results are therefore not in conflict with theirs, for our findings do not bear on the concept of secondary suggestibility.

Although we were alerted for evidence related to his distinction, we found nothing that permitted us to classify most subjects according to White's (1937) distinction between active and passive subjects.

Das (1958) has recently reported a factor analytical study of a scale of hynotic depth and reports finding a strong general factor and a second much weaker factor accounting for most hypnotic suggestibility. He used very few subjects, however, and such results as he found must be rather unstable.

USES OF A SUSCEPTIBILITY SCALE

In validating our scale as one measuring susceptibility we have attempted to show that it selects those who, when given further opportunities to experience hypnosis, are more likely to show the more varied phenomena associated with an established trance, in contrast to those who with further opportunities are less likely to show the phenomena of hypnosis. The distinction is in this respect one familiar in psychology between aptitude tests and achievement tests: an aptitude test predicts what can happen with further experience, while an achievement test indicates what profiting there has already been from experience.

At the same time, we are making use of the scores on our scale as criterion scores in the studying of correlates of hypnotic susceptibility. This uses the test scores as samples of hypnotic performance. It is entirely permissible to use the same set of scores as a criterion for one purpose and as themselves predictive scores for other purposes. Thus we use college grades as criteria by which to validate scholastic aptitude tests, but we also use college grades as predictive of graduate work.

To the extent that a susceptibility scale predicts those who are good candidates for further hypnosis, it is useful in discovering promising subjects. This is, however, by no means its only use. Another use is finding what degree of hypnotic susceptibility is needed for certain other purposes, as in the use of hypnosis in dentistry, obstetrics, or psychotherapy. Another use is to find how susceptibility is modified by changes in motivation such as occur in confronting childbirth, surgery, the pain of burns, and so on. For studies in which changes in susceptibility are to be investigated, the availability of equivalent alternate forms (Forms A and B) will prove to be of considerable service.

Other kinds of scales can be developed and are in process of construction in this laboratory.[5]

We have thus presented our susceptibility scale as but one kind of measure of hypnotic ability. Its primary purpose is to find out whether or not a given subject is likely to achieve a satisfactory trance with further hypnotic experience. The evidence at hand suggests that it is a fairly satisfactory measure for this purpose. It will prove to be of greater service when more norms are available from populations other than university students. To the extent that many experiments are done with university students, even the limited norms are of value.

Summary

1. Analyses were made of the responses of 124 college students (64 men and 60 women) to a newly developed scale of hypnotic susceptibility known as the Stanford Hypnotic Susceptibility Scale. The scale has been prepared in two highly similar forms (Form A and Form B), each yielding scores ranging from 0 to 12. All subjects were scored on both forms, one score being obtained on each of 2 days of hypnotic induction.

2. For the purposes of these analyses hypnotic susceptibility is defined as the number of responses representative of hypnosis yielded within the standard procedures of attempted induction and testing. As a sample of hypnotic phenomena the scale provides a *criterion* for personality studies; as an *aptitude test* it predicts the capacity to go on for more varied and complex hypnotic experiences.

3. The historical reports on susceptibility usually have referred to the greatest depth of hypnotic trance achieved under various methods of induction, and with repeated sessions.

While there is much variation, a summary of nineteenth century studies indicated a mean result of 9% refractory (nonsusceptible), 29% reaching a drowsy-light state, 36% moderately hypnotizable ("hypnotaxy"), and 26% reaching a deep or somnambulistic trance. A similar summary of investigations since 1930 indicated 22% refractory, 42% drowsy-light, 26% moderate, and 10% deep or somnambulistic. This order of disagreement does not seem at all surprising in view of the many uncertainties in this kind of attempted quantitative comparison.

Our own results, if divided in this fashion, would yield about 17% refractory, 35% drowsy-light, 25% moderate, and 23% deep or somnambulistic, thus falling very much in line with earlier studies.

4. The sample studied was compared with the total student body at Stanford, and with a larger sample of the introductory psychology course from which our subjects were drawn. The sample is moderately well representative of Stanford undergraduates.

5. The distribution of susceptibility turns out to be bimodal in 1958–59 as it was in

[5] Part II, as described earlier in this report, is undergoing extensive revision. Another kind of scale has been developed for special purposes (Weitzenhoffer & Sjoberg, 1961).

1957–58. A special study was made of artifacts that could lead to bimodality. These include: (*a*) high item intercorrelation for dichotomized items, and (*b*) a distribution of item difficulty that includes several items of nearly equal difficulty. Because both of these artifacts are present in our data, the bimodality must be viewed with suspicion. The presence of bimodality in the amnesia scores, however, cannot be attributed to these artifacts, and so leaves open the question of a "genuine" bimodality underlying these data.

6. Various analyses of item intercorrelations lead to the belief that the scale is essentially unidimensional. The individual items show high biserial *r*'s with the total score less that item. The Guttman scale is close to meeting the standard of 90% reproducibility; a factor analysis shows high saturation with a common factor.

7. Retest reliability of .83 and Kuder-Richardson reliability of .91 for the 2-day scores mean that the scales are satisfactory in establishing criteria for the studies of personality correlates that are now in progress. Some evidence of validity is provided by efforts to hypnotize by other methods selected subjects from the sample. The low scoring subjects proved refractory by other methods; the highest scoring subjects proved most able to go on to further hypnotic experiences.

REFERENCES

BARRY, H., MACKINNON, D. W., & MURRAY, H. A., JR. Studies on personality: A. Hypnotizability as a personality trait and its typological relations. *Hum. Biol.*, 1931, **13**, 1–36.

BEAUNIS, H. *Le somnambulisme provoqué.* Paris: Bailière, 1887.

BERNHEIM, H. *Hypnotisme, suggestion, psychothérapie: Etudes nouvelles.* Paris: Doin, 1891.

BRAID, J. *Neurypnology: Or the rationale of nervous sleep considered in relation to animal magnetism.* London: Churchill, 1843.

BRAMWELL, J. M. *Hypnotism.* (Originally published 1903) New York: Julian, 1956.

CHARCOT, J. M. Physiologie pathologique. Sur les divers états nerveux déterminés par l'hypnotization chez les hystériques, *CR Acad. Sci., Paris*, 1882, **94**, 403–405.

DAS, J. P. Factor analysis of a hypnotic scale. *Indian J. Psychol.*, 1958, **33**, 92–100.

DAVIS, L. W., & HUSBAND, R. W. A study of hypnotic susceptibility in relation to personality traits. *J. abnorm. soc. Psychol.*, 1931, **26**, 175–182.

EYSENCK, H. J., & FURNEAUX, W. D. Primary and secondary suggestibility: An experimental and statistical study. *J. exp. Psychol.*, 1945, **35**, 485–503.

FRIEDLANDER, J. W., & SARBIN, T. R. The depth of hypnosis. *J. abnorm. soc. Psychol.*, 1938, **33**, 281–294.

GILLES DE LA TOURETTE. *L'hypnotisme et les états analogues au point de vue médico-légal.* Paris: Plon, Nourrit, & Cie, 1889.

GOUGH, H. G. *Manual for the California Psychological Inventory.* Palo Alto, Calif.: Consulting Psychologists Press, 1957.

GUILFORD, J. P. *Fundamental statistics in psychology and education.* (2nd ed.) New York: McGraw-Hill, 1950.

GUTTMAN, L. The basis for scalogram analysis. In S. A. Stouffer and others (Eds.), *Measurement and prediction.* Princeton, N. J.: Princeton Univer. Press, 1950. Pp. 60–90.

HILGARD, E. R., & HOMMEL, L. S. Selective amnesia for events within hypnosis in relation to repression. *J. Pers.*, 1961, in press.

HILGARD, E. R., WEITZENHOFFER, A. M., & GOUGH, P. Individual differences in susceptibility to hypnosis. *Proc. Nat. Acad. Sci.*, 1958, **44**, 1255–1259.

HULL, C. L. *Hypnosis and suggestibility: An experimental approach.* New York: Appleton-Century, 1933.

LECRON, L., & BORDEAUX, J. *Hypnotism today.* New York: Grune & Stratton, 1947.

LIÉBEAULT, A. A. *Le sommeil provoqué et les états analogues.* Paris: Doin, 1889.

LOEWENFELD, L. *Der hypnotismus.* Wiesbaden: Bergmann, 1901.

MAYO, H. *Popular superstitions, and the truths contained therein, with an account of mesmerism.* Philadelphia: Lindsay & Blakiston, 1852.

MOORE, ROSEMARIE K. Relationship of susceptibility to hypnosis and susceptibility to social influence. Unpublished PhD dissertation, Stanford University, 1961.

PITRES, A. *Lectures cliniques sur l'hystérie et l'hypnotisme.* Vol. 2. Paris: Doin, 1891.

RICHER, P. *Études cliniques sur la grande hystérie ou hystéro-épilepsie.* Paris: Delahaye & Lecrosnier, 1885.

RICHET, C. *L'homme et l'intelligence.* Paris: Alcan, 1884.

SCHMIDKUNZ, H. *Psychologie der suggestion.* Stuttgart: Enke, 1892.

SCHMIDKUNZ, H. Zur statistik des hypnotismus. *Wien. med. Wschr.*, 1894, **23**, 1022–1024.

STUKÁT, K-G. *Suggestibility: A factorial and experimental analysis.* Stockholm: Almqvist & Wiksell, 1958.

WATKINS, J. G. *Hypnotherapy of the war neuroses.* New York: Ronald, 1949.

WEITZENHOFFER, A. M. Hypnotic susceptibility as related to masculinity-femininity. Unpublished PhD dissertation, University of Michigan, 1956.

WEITZENHOFFER, A. M. *General techniques of hypnotism.* New York: Grune & Stratton, 1957.

WEITZENHOFFER, A. M., & HILGARD, E. R. *Stanford Hypnotic Susceptibility Scale.* Palo Alto, Calif.: Consulting Psychologists Press, 1959.

WEITZENHOFFER, A. M., & SJOBERG, B. Suggestibility with and without "induction of hypnosis." *J. nerv. ment. Dis.*, 1961, **132**, 204–220.

WEITZENHOFFER, A. M., & WEITZENHOFFER, G. B. Sex, transference, and susceptibility to hypnosis. *Amer. J. clin. Hypnosis*, 1958, **1**, 15–24.

WHITE, M. M. The physical and mental traits of individuals susceptible to hypnosis. *J. abnorm. soc. Psychol.*, 1930, **25**, 293–298.

WHITE, R. W. Two types of hypnotic trance and their personality correlates. *J. Psychol.*, 1937, **3**, 265–277.

(Received July 16, 1960)